

CLAIMS

That which is claimed:

1. A method, comprising:

receiving a plurality of articles; and

identifying at least a first article as a shopping article.
2. The method of claim 1, wherein the first article is identified as a shopping article at least in part by identifying at least one price representation in a first article.
3. The method of claim 1, wherein the first article is identified as a shopping article at least in part by identifying at least one shopping character string in a link element or a form element of the first article.
4. The method of claim 1, further comprising providing an indication that the first article is a shopping article.
5. The method of claim 2, wherein the price representation is a currency symbol followed by a number followed by a period or comma followed by two single digit numbers.

6. The method of claim 3, wherein the character string comprises at least one of the group of character strings comprising add to cart, add to basket, add to shopping bag, update order, cart, basket, and checkout.
7. The method of claim 1, further comprising:
receiving a search query for an item;
associating the first article with the search query; and
identifying a first attribute associated with a first item relevant to the search query from the first article based at least in part on the search query.
8. The method of claim 7, further comprising identifying a second attribute associated with the first item based at least in part on the search query and the first attribute.
9. The method of claim 8, further comprising extracting the first attribute and the second attribute from the first article.
10. A method, comprising:
receiving a search query for an item;
identifying a first article associated with the search query;

identifying a first attribute associated with a first item relevant to the search query from the first article based at least in part on the search query.

11. The method of claim 10, further comprising identifying a second attribute associated with the first item based at least in part on the search query and the first attribute.

12. The method of claim 11, further comprising extracting the first attribute and the second attribute from the first article.

13. The method of claim 10, further comprising identifying a plurality of attributes associated with a plurality of items relevant to the search query from the first article.

14. The method of claim 11, wherein the identification of the first attribute and the second attribute is based at least in part on a structure of the first article.

15. The method of claim 10, wherein identifying the first attribute comprises determining a relationship between the first attribute and a query term.

16. The method of claim 15, wherein determining the relationship comprises determining a number of words between the first attribute and a query term.

17. The method of claim 15, wherein the first article has a tree structure and determining the relationship comprises determining a distance from the first attribute and a query term to a closest common ancestor.

18. The method of claim 15, wherein the first article has a tree structure and determining the relationship comprises determining a number of nodes in a smallest tree that contains both the first attribute and a query term.

19. The method of claim 15, wherein the first article has a tree structure and determining the relationship comprises determining a depth of a smallest tree in the tree structure containing both the first attribute and a query term.

20. The method of claim 11, wherein identifying the first attribute comprises determining a distance between the first attribute and the second attribute.

21. The method of claim 10, wherein the first attribute is a price for the first item.

22. The method of claim 21, wherein identifying the price comprises determining a price representation score.

23. The method of claim 21, wherein identifying the price comprises determining a font size of the price.

24. The method of claim 21, wherein identifying the price comprises determining a font face of the price.

25. The method of claim 21, wherein identifying the price comprises determining words immediately preceding the price.

26. The method of claim 11, wherein identifying the first attribute and the second attribute comprises determining global information associated with articles related to the first article.

27. The method of claim 11, wherein identifying the second attribute comprises determining a relationship between the second attribute and a query term.

28. The method of claim 27, wherein determining the relationship comprises determining a number of words between the second attribute and a query term.

29. The method of claim 27, wherein the first article has a tree structure and determining the relationship comprises determining a distance from the second attribute and a query term to a closest common ancestor.

30. The method of claim 27, wherein the first article has a tree structure and determining the relationship comprises determining a number of nodes in a smallest tree that contains both the second attribute and a query term.

31. The method of claim 27, wherein the first article has a tree structure and determining the relationship comprises determining a depth of a smallest tree in the tree structure containing both the second attribute and a query term

32. The method of claim 11, wherein identifying the second attribute comprises determining a distance between the second attribute and the first attribute.

33. The method of claim 11, wherein the second attribute is an image of the first item.

34. The method of claim 33, wherein identifying the image comprises determining an aspect ratio associated with the image.

35. The method of claim 33, wherein identifying the image comprises determining a frequency of occurrence value associated with the image.

36. The method of claim 11, wherein the identification of the first attribute and the second attribute is performed simultaneously.

37. A computer-readable medium containing program code, comprising:
program code for receiving a plurality of articles; and
program code for identifying the first article as a shopping article.

38. The computer-readable medium of claim 37, wherein the first article is identified as a shopping article at least in part by identifying at least one price representation in a first article.

39. The computer-readable medium of claim 37, wherein the first article is identified as a shopping article at least in part by identifying at least one shopping character string in a link element or a form element of the first article.

40. The computer-readable medium of claim 37, further comprises program code for providing an indication that the first article is a shopping article.

41. The computer-readable medium of claim 38, wherein the price representation is a currency symbol followed by a number followed by a period or comma followed by two single digit numbers.

42. The computer-readable medium of claim 35, wherein the character string comprises at least one of the group of character strings comprising add to cart, add to basket, add to shopping bag, update order, cart, basket, and checkout.

43. The computer-readable medium of claim 37, further comprising:
program code for receiving a search query for an item;
program code for associating the first article with the search query;
program code for identifying a first attribute associated with the item from the first article based at least in part on the search query.

44. The computer-readable medium of claim 43, further comprising program code for identifying a second attribute associated with the item based at least in part on the search query and the first attribute.

45. The computer-readable medium of claim 44, further comprising program code for extracting the first attribute and the second attribute from the first article.

46. A computer-readable medium containing program code, comprising:
- program code for receiving a search query for an item;
- program code for identifying a first article associated with the search query;
- program code for identifying a first attribute associated with a first item relevant to the search query from the first article based at least in part on the search query.
47. The computer-readable medium of claim 46, further comprising program code for identifying a second attribute associated with the first item based at least in part on the search query and the first attribute.
48. The computer-readable medium of claim 47, further comprising program code for extracting the first attribute and the second attribute from the first article.
49. The computer-readable medium of claim 46, further comprising program code for identifying a plurality of attributes associated with a plurality of items relevant to the search query from the first article.
50. The computer-readable medium of claim 47, wherein the identification of the first attribute and the second attribute is based at least in part on a structure of the first article.

51. The computer-readable medium of claim 46, wherein identifying the first attribute comprises program code for determining a relationship between the first attribute and a query term.

52. The computer-readable medium of claim 51, wherein determining the relationship comprises program code for determining a number of words between the first attribute and a query term.

53. The computer-readable medium of claim 51, wherein the first article has a tree structure and determining the relationship comprises program code for determining a distance from the first attribute and a query term to a closest common ancestor.

54. The computer-readable medium of claim 51, wherein the first article has a tree structure and determining the relationship comprises program code for determining a number of nodes in a smallest tree that contains both the first attribute and a query term.

55. The computer-readable medium of claim 51, wherein the first article has a tree structure and determining the relationship program code for comprises determining a

depth of a smallest tree in the tree structure containing both the first attribute and a query term

56. The computer-readable medium of claim 47, wherein identifying the first attribute comprises program code for determining a distance between the first attribute and the second attribute.

57. The computer-readable medium of claim 46, wherein the first attribute is a price for the first item.

58. The computer-readable medium of claim 57, wherein identifying the price comprises program code for determining a price representation score.

59. The computer-readable medium of claim 57, wherein identifying the price comprises program code for determining a font size of the price.

60. The computer-readable medium of claim 57, wherein identifying the price comprises program code for determining a font face of the price.

61. The computer-readable medium of claim 57, wherein identifying the price comprises program code for determining words immediately preceding the price.

62. The computer-readable medium of claim 47, wherein identifying the first attribute and the second attribute comprises program code for determining global information associated with articles related to the first article.

63. The computer-readable medium of claim 47, wherein identifying the second attribute comprises program code for determining a relationship between the second attribute and a query term.

64. The computer-readable medium of claim 63, wherein determining the relationship comprises program code for determining a number of words between the second attribute and a query term.

65. The computer-readable medium of claim 63, wherein the first article has a tree structure and determining the relationship comprises program code for determining a distance from the second attribute and a query term to a closest common ancestor.

66. The computer-readable medium of claim 63, wherein the first article has a tree structure and determining the relationship comprises program code for determining a number of nodes in a smallest tree that contains both the second attribute and a query term.

67. The computer-readable medium of claim 63, wherein the first article has a tree structure and determining the relationship comprises program code for determining a depth of a smallest tree in the tree structure containing both the second attribute and a query term

68. The computer-readable medium of claim 47, wherein identifying the second attribute comprises program code for determining a distance between the second attribute and the first attribute.

69. The computer-readable medium of claim 47, wherein the second attribute is an image of the first item.

70. The computer-readable medium of claim 69, wherein identifying the image comprises program code for determining an aspect ratio associated with the image.

71. The computer-readable medium of claim 69, wherein identifying the image comprises program code for determining a frequency of occurrence value associated with the image.

72. The computer-readable medium of claim 47, wherein the identification of the first attribute and the second attribute is performed simultaneously.

73. A method, comprising:

receiving a plurality of articles;

identifying at least a first article as a shopping article, wherein the first article is identified as a shopping article at least in part by identifying at least one price representation in a first article and at least in part by identifying at least one shopping character string in a link element or a form element of the first article;

receiving a search query for an item;

associating the first article with the search query;

identifying a price associated with a first item relevant to the search query from the first article based at least in part on the search query; and

identifying an image associated with the first item based at least in part on the search query and the price.